

MAREDATA: RED TEMÁTICA SOBRE DATOS DE INVESTIGACIÓN EN ABIERTO

ERNEST ABADAL, RAFAEL ALEIXANDRE, AGUSTÍ CANALS, ANTÒNIA
FERRER, IRENE GARRIGÓS, TONY HERNÁNDEZ, ALEXANDRE LÓPEZ-
BORRULL, JOSE N. MAZÓN, REMEDIOS MELERO, EVA MÉNDEZ,
CANDELA OLLÉ y FERNANDA PESET

RESUMEN: Se presentan los objetivos, las actividades y las líneas de trabajo de Maredata, una red temática dedicada a la gestión de datos de investigación y financiada por el Plan Estatal de I+D+I. Maredata ha sido creada por cinco grupos de investigación españoles con el propósito de difundir el interés por la gestión de datos de investigación entre la comunidad académica (investigadores, gestores de proyectos, bibliotecarios, etc.) y de generar nuevas líneas de investigación al respecto.

Palabras clave: Datos de investigación; redes temáticas; reutilización; Maredata.

ABSTRACT: We present the objectives, activities and research lines of Maredata, a thematic network funded by the Spanish National R+D Program. Maredata has been created by five research groups for spreading interest on the research data management among the academic community (not only researchers but also librarians and research managers) and for generating new lines of research.

Keywords: research data; reuse; academic networks; Maredata.

I. INTRODUCCIÓN

En estos momentos, la ciencia abierta (*Open Science*) es el marco de referencia para la investigación y la comunicación científica. Este término hace referencia a trabajar con una visión «abierta» (*open*) en todas las etapas o fases de la investigación científica, ya sea el diseño metodológico, la recogida de datos, la revisión por expertos, la publicación o la evaluación del impacto. Para dejar bien claro este nuevo enfoque, la Comisión Europea presenta en 2016 la *European Cloud Initiative*, que muestra cuál debe ser la hoja de ruta para conseguir la *European Open Science Cloud*, un entorno común de datos, *software*, estándares, experiencia y políticas que quiere ser una realidad en el año 2020.

Los datos científicos –base de cualquier investigación– son, sin ninguna duda, uno de los elementos clave de la ciencia abierta. Esto ya fue apuntado por la misma Comisión Europea con su plan piloto para los datos científicos de proyectos de investigación financiados por el programa Horizon2020, ya que obligaba a elaborar un plan de gestión de datos y a ponerlos en abierto en siete áreas del conocimiento (a partir de enero de 2017 se amplía este programa piloto para todos los proyectos nuevos).

Así pues, la importancia de los datos de investigación y de su gestión, así como el enfoque hacia su reutilización están fuera de toda duda. En los últimos años se ha ido incrementando este interés y todo hace suponer que se va a acentuar aún más en los próximos años, como quedó claro en la propuesta de objetivos para 2020 establecidos por un comité de expertos reunidos en el período en que los Países Bajos ejercía la Presidencia de la Unión Europea. En este documento se deja claro que después de la consolidación del acceso abierto a las publicaciones se tiene que priorizar la gestión de los datos de investigación, para los cuales el establecimiento de estándares e infraestructuras son puntos cruciales:

- «– Full open access for all scientific publications.
- A fundamentally new approach towards optimal reuse of research data. Data sharing and stewardship is the default approach for all publicly funded research. This requires definitions, standards and infrastructures». (Amsterdam, 2016).

Por otra parte, el valor de los datos no es solo científico. Su importancia para el avance económico y de la innovación fue también señalado por los comisarios europeos Moedas y Oettinger (Commissioner for Research, Science and Innovation y Commissioner for Digital Economy & Society, respectivamente) (Moedas, 2015) quienes declararon la intención de la Comisión Europea de trabajar por una economía basada en datos: «open science for a knowledge and data-driven economy» (Oettinger, 2015).

Una destacada representación de organizaciones e instituciones internacionales ya han elaborado documentos de políticas para asegurar la adecuada conservación y difusión de los datos científicos. Además de la Comisión Europea (2014; 2016) antes comentada, también se tiene que hacer referencia a Naciones Unidas (2014) con el informe «Mobilising the data revolution for sustainable development», la OCDE (2015) con «Making Open Science a reality», o la Liga Europea de Universidades de Investigación (LERU, 2013) con «LERU roadmap for research data» o el proyecto de investigación LEARN (2017), que ha elaborado un formulario para redactar políticas propias. Todos ellos conceden a este ámbito unas destacadas perspectivas de desarrollo hasta su plena consolidación.

Diversos son los aspectos relacionados con la gestión de datos que se están estudiando en estos momentos, ya sea el establecimiento de metadatos para la descripción de los ficheros de datos de investigación (Greenberg, 2010; Gómez, et al., 2016), las cuestiones relacionadas con el archivo en repositorios específicos que no están tan desarrollados como los destinados a publicaciones y revistas científicas (García-García, et al., 2015), así como conocer los hábitos y las opiniones de los investigadores mediante encuestas y entrevistas (Tenopir, 2011; 2015; Ferguson, 2014; Aleixandre, et al., 2015), entre otras cuestiones. En lo que respecta a monografías en español, se han publicado dos estudios que abordan esta cuestión de manera general (Couto, 2016) (Pezet & González, 2017) así como un informe de trabajo muy completo elaborado por un grupo de trabajo del proyecto Recolecta (Grupo, 2012).

El papel de la bibliotecas en el apoyo a la gestión de datos de investigación (metadatos, repositorios, planes de gestión, etc.) está siendo destacable y, en España, dispone de una notable consolidación. Podemos hacer referencia a diversos grupos de trabajo creados en Recolecta (Depósito y Gestión de datos en Acceso Abierto) –antes mencionado–, REBIUN (la línea 2 dedicada a Aprendizaje e Investigación), el CSUC (Grup de Treball de Suport a la Recerca), así como a diversos documentos elaborados por ellos con consejos y recomendaciones dirigidas a bibliotecas y a científicos (REBIUN, 2016; CSUC, 2017). También se tiene que destacar el repositorio e-cienciaDatos, creado por el consorcio Madroño para albergar los datos científicos obtenidos por los investigadores de las universidades públicas de la Comunidad de Madrid y la UNED.

2. SOBRE MAREDATA

Las redes temáticas tienen por objetivo agrupar las actividades de diversos grupos de investigación de un mismo ámbito para crear sinergias de

colaboración y generar nuevos proyectos. En este contexto se constituyó la red temática Maredata en 2015, que tiene por objetivo general aglutinar y consolidar la colaboración entre grupos de investigación españoles centrados en el estudio de los datos científicos y también establecer relación con aquellos sectores y disciplinas interesados en esta temática. El proyecto fue financiado por el Plan Estatal I+D+I (proyecto CSO2015-71867-REDT) en su convocatoria de 2015 y finalizará en septiembre de 2018.

2.1. MIEMBROS

Los miembros de la red pertenecen a siete instituciones distintas: Instituto de Alimentación y Tecnología de Alimentos (CSIC), INGENIO (CSIC-Universitat de València), Universidad de Alicante, Universitat de Barcelona, Universidad Carlos III de Madrid, Universitat Oberta de Catalunya y Universidad Politécnica de Valencia, que participan en cinco proyectos de investigación financiados por el Plan Estatal.

- Acceso abierto a la ciencia (UB y IATA): <http://accesoabierto.net>
- Curator-e: Custodia y gestión digital de datos de investigación (UC3M): <http://www.curatore.es>
- Knowledge and Information Management in Organizations (KIMO) (UOC): <http://kimo.uoc.edu>
- Datasea: Datos abiertos de investigación (INGENIO y UPV): <http://www.datasea.es>
- WaKe (Web and Knowledge) (UA): <http://rua.ua.es/dspace/handle/10045/22407?locale=en>

2.2 OBJETIVOS ESPECÍFICOS

- a) Coordinar las actuaciones y líneas de investigación de los distintos grupos.

Se están analizando las actuaciones y líneas de investigación de cada uno de los grupos de investigación para promover proyectos colaborativos y provocar sinergias entre ellos.

- b) Detectar agentes interesados en el tratamiento y difusión de los datos de investigación.

A través de eventos conjuntos se está contactando no solo con investigadores de todas las disciplinas, sino también con bibliotecarios y otros profesionales de la información, empresas de servicios, autoridades

académicas y gestores de proyectos de investigación. Es importante la colaboración de todos estos agentes para que la ciencia española avance en la gestión de datos de investigación al ritmo que se está demandando a nivel internacional.

- c) Identificar y contactar con los grupos de investigación que generan datos de científicos en proyectos de H2020.

Estos grupos de investigación son los más concienciados sobre la necesidad del tratamiento de los datos científicos. Se va a determinar cuáles son estos grupos (inicialmente en Cataluña, Madrid y Valencia) para conocer con detalle sus intereses y problemas.

- d) Facilitar la generación de nuevas líneas de investigación.

El trabajo coordinado de la red permitirá la generación de nuevas propuestas científicas en torno a *Open Data* y *Open Science*, ya sean conjuntas o en colaboración, que serán de mucha utilidad e interés para avanzar en las cuestiones relativas a la gestión de los datos científicos.

- e) Facilitar y fomentar la internacionalización de los grupos participantes.

La creación de la red facilitará a sus miembros participar en los foros internacionales sobre datos abiertos, datos de investigación, acceso abierto, *open science*, etc. Por otra parte, los contactos internacionales van a facilitar también posibles participaciones en proyectos europeos del programa H2020.

- f) Elaborar recomendaciones sobre gestión de los datos de investigación.

Como resultado de la identificación de los grupos de investigación generadores de datos científicos, así como de los agentes interesados en ellos (empresas, bibliotecas académicas, gestores de proyectos, etc.), se conocerán con detalle las necesidades y problemáticas y se podrán elaborar unas recomendaciones sobre cómo se puede proceder en este ámbito.

2.3. SEMINARIOS Y CONFERENCIAS

Durante los años 2016 y 2017 se han organizado cuatro eventos centrados en la gestión de los datos de investigación

- a) Seminario sobre gestión de datos de investigación (Barcelona: Universitat de Barcelona, 19 de enero de 2016).

Se estructuró en torno a cuatro grandes bloques temáticos: las políticas, los investigadores, las infraestructuras y los aspectos legales. Contó con la presencia de nueve ponentes y más de un centenar de asistentes.

- b) Taller «¿Qué sabemos de los datos de investigación?», XV Workshop de REBIUN (Castelló: Universitat Jaume I, 29 de septiembre de 2016).

Se organizó un taller práctico a partir de las preguntas básicas relacionadas con los datos científicos.

- c) Taller «Datos abiertos de investigación o Investigación con Datos Abiertos», Pre-evento IODC (International Open Data Conference) (Madrid: Universidad Carlos III, 5 de octubre de 2016).

El propósito era discutir y analizar los dos papeles principales en el escenario de los datos abiertos de investigación: los que producen *open data* como resultado de su investigación, y los que producen datos abiertos gubernamentales susceptibles de utilizarse para la investigación. Se contó con la participación de 16 ponentes y unos treinta asistentes.

- d) Towards RDA Iberia. Evento asociado al RDA Ninth Plenary Meeting (Barcelona, 4 de abril 2017).

Este *workshop* fue preparado para fomentar la creación de RDA Iberia y tuvo como prioridad colaborar en las políticas y estándares relacionados con la apertura y reutilización de los datos de investigación, una parte fundamental de la agenda de la *Open Science* en Europa. Contó con más de 90 asistentes.

Por otra parte, en lo que se refiere a la participación, hemos estado presentes, con pósters o comunicaciones sobre la red Maredata en los eventos siguientes:

- Melero, R.; López-Borrull, A.; Ferrer-Sapena, A.; Abadal, E.; Peset, F.; Aleixandre, R.; Olle, C.; Hernandez, T.; Méndez, E.; Canals, A. (2016). «Maredata: Red española sobre datos de investigación en abierto». En: CRECS 2016 (Barcelona, 5-6 mayo 2016).
- Ferrer, A.; Aleixandre, R. (2016). «Los científicos ante los datos abiertos. Red Maredata». En: Seminario Universidad de Alicante.
- Grupo Maredata. (2017). «Maredata: métricas de una red temática». En: JEID Fesabid 2017 (Pamplona, 25-26 mayo 2017).
- López-Borrull, A.; Melero, R.; Ollé, C.; Peset, F. (Maredata) (2017). «Diez aspectos a considerar por parte de las revistas científicas en ciencias sociales antes de diseñar una política de datos». En: CRECS 2017 (Cuenca, mayo 2017).

2.4. DIFUSIÓN GENERAL

Se ha creado el portal Maredata (<http://maredata.net/>), que tiene por objetivo difundir la actividad de la red entre la comunidad académica. Dispone de tres herramientas básicas de comunicación: un canal de noticias al cual es posible suscribirse (cuenta en estos momentos con unos 150 suscritos), una sección dedicada a los eventos (incluye todo el material difundido en los seminarios organizados por la red temática) y otra con la producción científica derivada del proyecto.

La presencia en medios de comunicación también se considera especialmente relevante para conseguir incrementar la visibilidad de la red. Precisamente en relación a esta cuestión se quiere destacar que los miembros de la red están siendo activos difundiendo el interés por los datos de investigación entre el gran público a través de los medios de comunicación, destacando entre ellos la presencia en TVE¹.

2.5. VALORACIÓN PROVISIONAL

Hasta estos momentos, por tanto, el ritmo de consecución de los objetivos planteados permite suponer que se logrará cumplir con todos ellos al finalizar el proyecto. De todas formas, a partir de la valoración del trabajo realizado hasta ahora, podemos destacar cuatro aspectos en los cuales se ha avanzado de manera notable:

a) Internacionalización.

Es significativo que la mitad de los eventos organizados han sido llevados a cabo en el marco de congresos internacionales, ya sea IODC (International Open Data Conference) o el *Plenary Meeting* de RDA (Research Data Alliance). Se tiene que destacar, además, que RDA creada en 2013 por la Comisión Europea, la National Science Foundation (USA) y el Departamento de Innovación (Australia) es la principal organización internacional del ámbito y tiene por objetivo construir la infraestructura social y técnica que permita el intercambio abierto de datos. A día de hoy cuenta con más de 5.500 miembros individuales de 115 países.

¹ Se puede consultar en: <http://www.rtve.es/noticias/20160312/maredata-red-big-data-impulsada-csic-seis-universidades-espanolas-para-investigacion-abierto/1317764.shtml>.

b) Interrelación con otros ámbitos temáticos.

El proyecto nace de cinco grupos (y siete organizaciones) procedentes, en la mayoría de los casos, del ámbito de la Biblioteconomía y Documentación. A partir de las actividades relacionadas anteriormente se ha podido contactar con grupos de investigación, entidades, etc. vinculados a muchas otras disciplinas (ya sea la Química, la Agricultura, la Informática, etc.) o entornos (Barcelona SuperComputing Center, REBIUN, etc.). Con ellos hemos organizado algunos de los eventos y establecido relaciones que vamos a explotar y desarrollar en el futuro.

c) Creación de comunidad.

Como consecuencia de la interrelación antes descrita, se van a poder tejer alianzas con grupos y sectores que servirán para fundamentar el apoyo a la gestión de datos de investigación en España. En este sentido, destaca el desarrollo y consolidación de RDA Iberia como grupo geográfico dentro de RDA. Su objetivo es convertirse en el marco de encuentro y discusión para la gestión de datos de investigación en España y Portugal, tanto en los niveles políticos como técnicos. Así también, sobresale la colaboración con REBIUN, que no podemos olvidar que es la sectorial de bibliotecas de la CRUE y que, por tanto, tiene una destacada influencia en las universidades.

d) Detección de actores.

También se ha podido comprobar la multiplicidad de actores y *stakeholders* interesados en la gestión de datos de investigación, cada uno de ellos con agendas y necesidades muy diferentes. Así pues, investigadores, universidades, bibliotecas y revistas científicas tienen necesidades de formación e información distinta, pero es importante su interrelación porque disponen de tiempos encadenados.

3. BIBLIOGRAFÍA

- ALEIXANDRE-BENAVENT, R., VIDAL-INFER, A., ALONSO-ARROYO, A., FERRER-SAPENA, A., PESET, F., GARCÍA GARCÍA, A. (2015). Gestión de los datos brutos de investigación en los investigadores españoles en ciencias de la salud. *Trauma*, 26, 1, 73-81. <http://www.mapfre.com/fundacion/html/revistas/trauma/v26n1/articulo-11.html>.
- AMSTERDAM CALL FOR ACTION ON OPEN SCIENCE. (2016). Netherlands EU Presidency. <http://www.openaccess.nl/sites/www.openaccess.nl/files/documenten/amsterdam-call-for-action-on-open-science.pdf>
- COMISIÓN EUROPEA. (2014). Comunicación de la Comisión al Parlamento Europeo, al Consejo, al Comité Económico y Social Europeo y al Comité de las regio-

- nes. Hacia una economía de los datos próspera. COM (2014). 442 final. http://datos.gob.es/sites/default/files/com_2014_442_fr_communication_from_commission_to_inst_es_v2_pi_773238_4.pdf
- COMISIÓN EUROPEA. (2016). Comunicación de la Comisión al Parlamento Europeo, al Consejo, al Comité Económico y Social Europeo y al Comité de las Regiones. European cloud initiative – Building a competitive data and knowledge economy in Europe. COM (2016). 178 final. <http://eur-lex.europa.eu/legal-content/ES/TXT/PDF/?uri=COM:2016:178:FIN&rid=2>
- COUTO, F. (2016). *Gestión de datos de investigación*. Barcelona: Editorial UOC.
- CSUC (2017). Recomanacions per seleccionar un repositori per al dipòsit de dades de recerca: Versió 3. <http://hdl.handle.net/2072/284974>
- FERGUSON, L. (2014). How and why researchers share data (and why they don't). *Wiley Exchanges: Discover the future of research*. <https://hub.wiley.com/community/exchanges/discover/blog/2014/11/03/how-and-why-researchers-share-data-and-why-they-dont?referrer=exchanges>
- GREENBERG, J. (2010). Metadata for Scientific Data: Historical Considerations, Current Practice, and Prospects. *Journal of library metadata*, 10, 2-3, 75-78. <http://dx.doi.org/10.1080/19386389.2010.520262>
- HERNÁNDEZ-PÉREZ, T. (2016). «En la era de la web de los datos: primero datos abiertos, después datos masivos». *El profesional de la información*, 25, 4, 517-525. <http://dx.doi.org/10.3145/epi.2016.jul.01>
- LEARN. (2017). *Toolkit of best practice for research data management*. http://learn-rdm.eu/wp-content/uploads/RDMToolkit_rev06-17.pdf
- LERU. (2013). LERU roadmap for research data. Advice paper n. 14. http://www.leru.org/files/publications/API4_LERU_Roadmap_for_Research_data_final.pdf
- GARCÍA GARCÍA, A.; LÓPEZ BORRULL, A., & PESET, F. (2015). Data journals: eclosión de nuevas revistas especializadas en datos. *El profesional de la información*, 24, 6, 845-854
- GÓMEZ, N.-D.; MÉNDEZ, E.; HERNÁNDEZ-PÉREZ, T. (2016). Social sciences and humanities research data and metadata: A perspective from thematic data repositories. *El profesional de la información*, 25(4), 545-555. <http://dx.doi.org/10.3145/epi.2016.jul.04>
- GRUPO DE TRABAJO DE «Depósito y Gestión de datos en Acceso Abierto» del proyecto Recolecta. (2012). *La conservación y reutilización de los datos científicos en España. Informe del grupo de trabajo de buenas prácticas*. Madrid: FECYT. <http://eprints.rclis.org/21007/>
- NACIONES UNIDAS. (2014). *A world that counts: Mobilising the data revolution for sustainable development*. <http://www.undatarevolution.org/wp-content/uploads/2014/11/A-World-That-Counts.pdf>
- MELERO, R. & HERNÁNDEZ-SAN-MIGUEL, J. (2014). Acceso abierto a los datos de investigación, una vía hacia la colaboración científica. *Revista Española de Documentación Científica*, 37(4), eo66. <http://dx.doi.org/10.3989/redc.2014.4.1154>
- MOEDAS, C. (2015). Open Innovation, Open Science, Open to the World. In A new start for Europe: Opening up to an ERA of Innovation Conference SPEECH/15/5243 http://europa.eu/rapid/press-release_SPEECH-15-5243_en.htm

- OECD. (2015). Making Open Science a Reality. *OECD Science, Technology and Industry Policy Papers*, 25. <http://dx.doi.org/10.1787/5jrs2f963zsi-en>
- OETTINGER, G. H. (2015). Open science for a knowledge and data-driven economy. http://ec.europa.eu/commission/2014-2019/oettinger/blog/open-science-knowledge-and-data-driven-economy_en
- PESET, F. & GONZÁLEZ, L.-M. (2017). *Ciencia abierta y gestión de datos de investigación*. Gijón: TREA.
- REBIUN. (2016). Infografías 2016: documentación generada por la línea 2. <http://www.rebiun.org/documentos/Paginas/Documentaci%C3%B3n-generada-por-la-L%C3%ADnea-2.aspx>
- TENOPIR, C. et al. (2011). Data sharing by scientists: practices and perceptions. *PLOS ONE*, 6, 6, e21101. <http://dx.doi.org/10.1371/journal.pone.0021101>
- TENOPIR, C. et al (2015). Changes in data sharing and data reuse practices and perceptions among scientists worldwide. *PLOS ONE*, 10, 8. <http://dx.doi.org/10.1371/journal.pone.0134826>.